

**Propuesta metodológica para el estudio del lenguaje de la protesta y su transmisión a las redes sociales****A methodological proposal for the study of the protest language and its transmission to the social networks****Resumen**

El importante papel de las redes sociales virtuales en la proposición y coordinación de recientes movimientos de protesta exige nuevas propuestas metodológicas. Las características de la comunicación en red y su relación con las diversas formas de expresión visual del descontento sugieren la necesidad de establecer pautas que ofrezcan tanto un alto nivel descriptivo de los corpus como una sistematización estadística que permita operar eficientemente con los datos. La metodología analítica propuesta pretende ofrecer una guía en torno a cinco niveles de análisis aplicables a cualquier corpus escrito: fonético, morfosintáctico, semántico, léxico y retórico. A partir del establecimiento de las características dominantes de cada nivel a través del cotejo estadístico, se generará un perfil del conjunto del corpus.

**Palabras Clave**

Corpus, lingüística, metodología, movimientos sociales, redes sociales, Twitter.

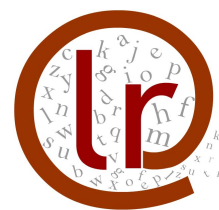
**Abstract**

The major role of the social networks on the proposal and coordination of the recent protest movements require new methodological proposals. The features of the network communication and their relations with the different ways of visual expression of the dissatisfaction suggest the need to set down standards which offer both a high descriptive level and a statistical systematization in order to operate efficiently with data. The proposed analytic methodology tries to offer a guide around five analytic categories applicable to any written corpus: phonetic, morph syntactic, semantic, lexical and rhetoric. From the establishment of the dominant characteristics of each category through the statistical crosschecking, a joint profile of the corpus will be generated.

**Key words**

Corpus, linguistics, methodology, social movements, social networks, Twitter.

Fecha de recepción: 04/04/2014 - Fecha de aceptación: 05/05/2014 – Fecha de publicación: 25/05/2014



## 1. Introducción

En los últimos tres años, las posibilidades de conexión que ofrecen las nuevas redes sociales virtuales han generado sensibles modificaciones en la organización de movimientos de protesta. Como reflejan Della Porta y Diani (2011: 160), “las redes sociales crean con frecuencia oportunidades para transformar en acción lo que hasta entonces era sólo una predisposición”. No parece por tanto una coincidencia la proliferación de movimientos como la Primavera Árabe en el norte de África, el movimiento 15M en España o sus versiones exportadas, como el *Occupy* norteamericano. Pese a estar viviendo una coincidencia temporal entre un salto tecnológico en el ámbito de la comunicación y la creación de movimientos sociales no exclusivamente basados en el entorno físico, expertos como Tilly y Wood (2009: 194) aconsejan:

“evitar el determinismo tecnológico; reconocer que la mayoría de los nuevos rasgos que presentan los movimientos sociales son el resultado de los cambios acaecidos en sus respectivos contextos sociales y políticos, y no el de las innovaciones tecnológicas como tales”.

Este escenario de proliferación de la protesta social se antoja fruto de una dualidad circunstancial, mezcla de un contexto socioeconómico de crisis persistente y de una fase de desarrollo exponencial de las vías de comunicación virtual. No obstante, esta motivación dual lleva camino de provocar, como afirma Yus (2010: 13), “un proceso de hibridación comunitaria físico-virtual, según la cual en una misma persona ‘conviven’ diferentes formas de intersubjetividad en redes personales de interacción”.



Imagen 1.1: Pancarta con el lema “Pienso luego estorbo”.  
Fuente: Martín Corvillo (2013)

## Propuesta metodológica para el estudio del lenguaje de la protesta y su transmisión a las redes sociales

En ese proceso de hibridación entre lo físico y lo virtual, los procesos de comunicación entre individuos que comparten total o parcialmente ambas esferas pueden arrojar evidencias de la estrecha influencia mutua de los fenómenos comunicativos que se dan tanto en lo real como en lo virtual. La constatación de que pancartas como la dispuesta en la Imagen 1.1 trascienden a la red y por tanto aumentan su potencial de difusión, como vemos en la Imagen 2.2, sugiere la presencia de una interrelación descriptible y de algún modo cuantificable.



**Imagen 1.2:** Tweet con el lema “Pienso luego estorbo”  
Fuente: Martín Corvillo (2013)

No obstante, no todos los lemas de protesta plasmados en pancartas llegan a alcanzar la esfera virtual. El porqué del éxito en términos de trascendencia de ciertos lemas y del fracaso del resto puede fundamentarse en cuestiones colindantes con la percepción colectiva y el efecto contagio (Torres Nabel, 2009: 68):

“Una primera hipótesis al respecto explica el como a partir de un tópico discutible y sobre todo un tópico con cierto grado de popularidad, en este caso dada por la misma aplicación, determinándolo como “tema en boga” (*trend topic*), se produce un efecto de contagio en los usuarios, al verse influenciados a participar de dicho tema hasta volverlo un acontecimiento; evidentemente esto no pasa con todos los *trend topics*, solo algunos logran tener cierto grado de consecuencias.”

No obstante, y pese a que resulte obvia la influencia del colectivo en decisiones particulares como la replicación o transmisión de un lema, existirá una serie de factores perceptivos o cognitivos –factores subjetivos y por tanto individuales– que condicionarán el atractivo que cierto slogan pueda tener. Estos factores parecen, de acuerdo con Yus (2010: 85), estrechamente relacionados con la satisfacción de necesidades cognitivas, dado que los receptores de lemas y usuarios de la red tratarán de obtener aquella información que estén buscando “sin un esfuerzo excesivo de procesamiento, para lo cual tenderán a usar los enlaces de forma adecuada y tratarán de mantener la necesaria coherencia entre los textos enlazados”. Una circunstancia que, a posteriori, puede denotar una búsqueda intencional de la maximización del principio de relevancia en los creadores de slogans (Sperber y Wilson, 2004: 244):

“La tendencia cognitiva universal a maximizar la relevancia hace posible que, al menos en cierto grado, se puedan predecir y manipular los estados mentales de los demás. Si yo soy consciente de que usted tiene tendencia a escoger los estímulos más relevantes entre aquellos de los que dispone en su entorno y a procesarlos de tal modo que se maximice su relevancia, estoy entonces capacitado para producir un estímulo que con toda probabilidad atraerá su atención”.

Dado que la esfera virtual permite acudir a posteriori a los datos vertidos en la misma, parece admisible considerar que estos flujos de información predominantemente persuasiva pueden ser estudiados con el objetivo de determinar qué elementos presentes en los lemas generan una propensión a la trascendencia en la red. Con este objetivo, este artículo pretende plantear una metodología de investigación que ofrezca dos utilidades a la comunidad científica:

- a) Un alto nivel descriptivo, teniendo en cuenta los aspectos estructurales pero también estilísticos que condicionan la percepción de un slogan.
- b) La posibilidad de establecer una cuantificación numérica de cuantos más parámetros mejor, en aras de poder emplear la estadística como instrumento de interpretación.

Para ello, a continuación se ofrece una descripción del proceso de cuantificación y análisis seguido en la tesis doctoral *15M: Análisis de la entropía comunicativa* (Martín Corvillo, 2013)<sup>1</sup>.

## 2. La metodología propuesta

La metodología propuesta consta de siete fases:

1. Captación del corpus y etiquetaje de sus elementos.
2. Establecimiento de cinco niveles de estudio (fonético, morfosintáctico, semántico, léxico y retórico) y las variables de discriminación propias de cada nivel.
3. Marcación en función de variables de discriminación y resumen estadístico.
4. Establecimiento del perfil estadístico del conjunto del corpus y discriminación de las características dominantes.
5. Elaboración del subcorpus a través de la búsqueda de *tweets* mediante un buscador *online*.
6. Establecimiento del perfil estadístico del conjunto del subcorpus y discriminación de las características dominantes.
7. Contraste entre los perfiles estadísticos conjuntos de corpus y subcorpus.

<sup>1</sup> En la misma se disponen datos que no pretenden ser concluyentes en el contexto de este artículo –no así en el del citado trabajo doctoral, defendido satisfactoriamente en el año 2013- pero sí ilustrativos de una metodología apta para el estudio lingüístico de fenómenos complejos.

Dichas siete fases de estudio son descritas a continuación, incluyendo como ejemplos ilustrativos capturas y resultados obtenidos en el proceso de investigación remitido en varias ocasiones a lo largo de la introducción de este artículo.

## 2.1 Captación del corpus y etiquetaje de sus elementos

El corpus es captado fotográficamente durante la fase previa de trabajo de campo, a la que es conveniente asignar un periodo de tiempo a priori para evitar sobredimensionar el mismo y a su vez perder la referencia contextual: este tipo de fenómeno comunicativo es fruto de un clima socioeconómico concreto y, de alargarse el periodo de recogida, el contexto puede variar drásticamente, alterando las circunstancias referenciales y temáticas, generándose lemas distantes de los originales en dichos aspectos. En el caso que nos ocupa, el corpus fue captado durante el intervalo de un mes, entre el 16 de mayo y el 15 de junio de 2011.

A continuación, el corpus es transcrito ordenando alfabéticamente sus elementos en un sistema de bases de datos, como bien puedan ser Microsoft Access, Microsoft Excel o sus programas análogos de código abierto. Posteriormente, se otorga una referencia numérica que siempre presente el mismo número de dígitos en función del número total de elementos del corpus para facilitar cualquier tipo de búsqueda. En el caso del corpus que nos ocupa, el número total de lemas empleados fue 620, por lo que todas las etiquetas presentaron números de tres cifras, comenzando por 001. Los archivos fotográficos fueron almacenados tras renombrarse empleando los números asignados.

Teniendo en cuenta que en el punto 5 del proceso de investigación se discriminará un subcorpus a partir del conjunto inicial, es esencial en este proceso metodológico que los elementos del corpus original mantengan siempre su numeración con independencia de la evidente disminución a posteriori en el número de elementos. Esto prevendrá la confusión entre lemas y posibilitará un acceso ágil a los mismos tanto al analizar los resultados obtenidos en el análisis del corpus como a la hora de revisar las peculiaridades del subcorpus.

## 2.2 Establecimiento de cinco niveles de estudio y sus variables de discriminación

Dado que el corpus es escrito en origen y en tanto trascenderá de manera también escrita aun pudiendo haber adiciones de apoyo como imágenes, será analizado de acuerdo a cinco niveles de análisis factible en este soporte. Cuestiones como el uso de imagen, el análisis tipográfico o las peculiaridades cromáticas quedan fuera de los niveles de estudio. El criterio empleado se basa en que, dado que se busca realizar un análisis tanto en el soporte visual como en el escrito virtual, el análisis de las manifestaciones debe centrarse en aquellas características perdurables en el tránsito de una a otra superficie.

REF	LEMA	i!	¿?	i! + ¿?
001	A la puta calle nos dijeron y eso hacemos			
002	A pesar de vuestra soberbia se os tratará con dignidad			
003	A por ellos, que son pocos y cobardes			
004	¡A por ellos! Que son pocos y cobardes. Salud y república.	1		
005	¡A por la tercera ya!	1		
006	A ser libres.			

Imagen 2.1A: Ejemplo de discriminación positiva o marcación (lemas marcados)

Fuente: Martín Corvillo (2013)

Dos de los niveles de análisis presentan a su vez subniveles que posibilitan adecuar la complejidad de cada uno de los mismos –en este caso, el morfosintáctico y el semántico- a la exigencia de llevar a cabo un análisis estadístico con un nivel de detalle lo más elevado posible. Por ello, dentro de cada estrato analítico se incluye una serie de variables detectables propias de la naturaleza de cada nivel. Estas variables serán llamadas a partir de ahora de discriminación positiva o marcación, y su registro se efectúa a través de un procedimiento simple y útil a efectos de contabilización: contabilizando con un “1” en la celda pertinente (ver Imagen 2.1A) para disponer del sumatorio total al final de la misma hoja de conteo (Imagen 2.2B).

A	LEMA	i!	¿?	i! + ¿?
613	#Enjoy the #acampadavalencia			
614	#Estamoshartos			
615	# Hashtag los huevos!	1		
616	# HashtagYa!	1		
617	# Revolución se escribe con hashtag			
618	#spanishrevolution			
619	#Tienes voz			
620	#tomalaplaza			
	<b>TOTALES</b>	<b>59</b>	<b>22</b>	<b>1</b>

Imagen 2.1B: Ejemplo de discriminación positiva o marcación (sumatorio de marcaciones)

Fuente: Martín Corvillo (2013)

En el caso particular del nivel léxico, donde lo que se estudia es la extensión total de los lemas, el proceso no es de marcación sino de contabilización. Por tanto, los guarismos expresados no implican presencia de ningún elemento denotativo sino sencillamente su valor numérico, sea en palabras o caracteres como ilustra la Imagen 2.2A.



## Propuesta metodológica para el estudio del lenguaje de la protesta y su transmisión a las redes sociales

REF	LEMA	PALABRAS	CARACTERES
001	A la puta calle nos dijeron y eso hacemos	9	42
002	A pesar de vuestra soberbia se os tratará con dignidad	10	54
003	A por ellos, que son pocos y cobardes	8	37
004	¡A por ellos! Que son pocos y cobardes. Salud y república.	11	59
005	¡A por la tercera ya!	5	22
006	A ser libres.	3	13
007	Abre los ojos. Estamos aquí.	5	29
008	Abuelo esto va x ti	5	19
009	Aceptamos todo tipo de donaciones excepto: dinero / alcohol / trajes	11	69

Imagen 2.2A: Ejemplo de contabilización en elementos del corpus (palabras y caracteres).

Fuente: Martín Corvillo (2013)

Evidentemente, en casos de contabilización de elementos en un corpus no existe una sección del corpus a discriminar como marcada o no marcada, por lo que debe aprovecharse, como ilustra la Imagen 2.2B, para disponer un valor estadístico que interese al investigador, como el promedio. No obstante, el promedio no es el único valor estadístico susceptible de ser empleado en un proceso descriptivo como el que nos ocupa, como se explica en el punto 2.3 correspondiente a los sistemas de disposición estadística.

613	#Enjoy the #acampadavalencia	3	28
614	#Estamoshartos	1	14
615	# Hashtag los huevos!	4	21
616	# HashtagYa!	2	12
617	# Revolución se escribe con hashtag	6	35
618	#spanishrevolution	1	18
619	#Tienes voz	2	11
620	#tomalaplaza	1	12
	<b>PROMEDIO</b>	<b>6,830645161</b>	<b>38,67580645</b>

Imagen 2.2B: Ejemplo de contabilización en elementos del corpus y disposición de promedios.

Fuente: Martín Corvillo (2013)

Una vez explicado el sencillo sistema de marcación y contabilización de esta propuesta metodológica, se ofrece a continuación el esquema de la estructura de niveles y subniveles de análisis, y las categorías de discriminación o marcación propias de cada uno de ellos. Dado que lo excepcional por escaso en el ejemplo que nos ocupa es la contabilización, estas aparecen expresamente señaladas en el esquema, siendo el resto de categorías no señaladas ejemplos de discriminación positiva o marcación:

## 1. Nivel fonético

- Presencia de exclamación (“¡Únete madero, tú también eres obrero!”)
- Presencia de interrogación (“¿Cri\$í\$? ¿Qu€ cri\$í\$?”)
- Presencia combinada de exclamación e interrogación (“¿Demócrata? ¡No te calles! A la calle”)

## 2. Nivel morfosintáctico

**Propuesta metodológica para el estudio del lenguaje de la protesta y su transmisión a las redes sociales**

## a. Subnivel temporal

- Presencia de formas verbales de pasado ("La democracia ha durado lo que una demo")
- Presencia de formas verbales de presente ("A por ellos, que son pocos y cobardes. Salud y república")
- Presencia de formas verbales de futuro ("A pesar de vuesdtra soberbia se os tratará con dignidad")
- Presencia de formas verbales no temporales ("A ser libres")
- Presencia de varias formas verbales de distintos tiempos ("A la puta calle nos dijeron y eso hacemos").
- Ausencia de verbo ("Semáforo rojo al tráfico de influencias").

## b. Subnivel sintáctico

- Expresión ("De un pasado gris a un presente negro")
- Construcción simple ("Cualquier noche puede salir el sol")
- Construcción coordinada ("La barricada cierra la calle pero abre el camino").
- Construcción subordinada ("Me creeré la crisis cuando los ricos se suiciden en masa").
- Construcción yuxtapuesta ("Los ricos están ricos: ¡cómetelos!")
- Unidad compleja (sucesión de construcciones independientes) ("Of the people. By the people. For the people")

## 3. Nivel semántico

## a. Subnivel de frecuencias

- Contabilización del número de apariciones de vocablos en el corpus

## b. Subnivel temático

- Referencia a economía ("Saca la pasta del banco")
- Referencia a legislación o política ("Si no es social no es democracia")
- Referencia a TIC ("#Hashtag Ya")

## c. Subnivel léxico-semántico

- Empleo de expresión parasitaria o calco ("Manos arriba, esto es un contrato")

## 4. Nivel léxico

- Contabilización de número de palabras
- Contabilización de número de caracteres

## 5. Nivel retórico

- Empleo de figuras retórica por adición o supresión ("Digamos no, negaos")
- Empleo de figuras retóricas por repetición ("Democracia 1.0, obsoleta. Democracia 2.0, instalándose")
- Empleo de figuras retóricas por combinación ("D'hont: si es democracia participativa, déjame participar")
- Empleo de tropos ("Mercaderes de nuestro templo")





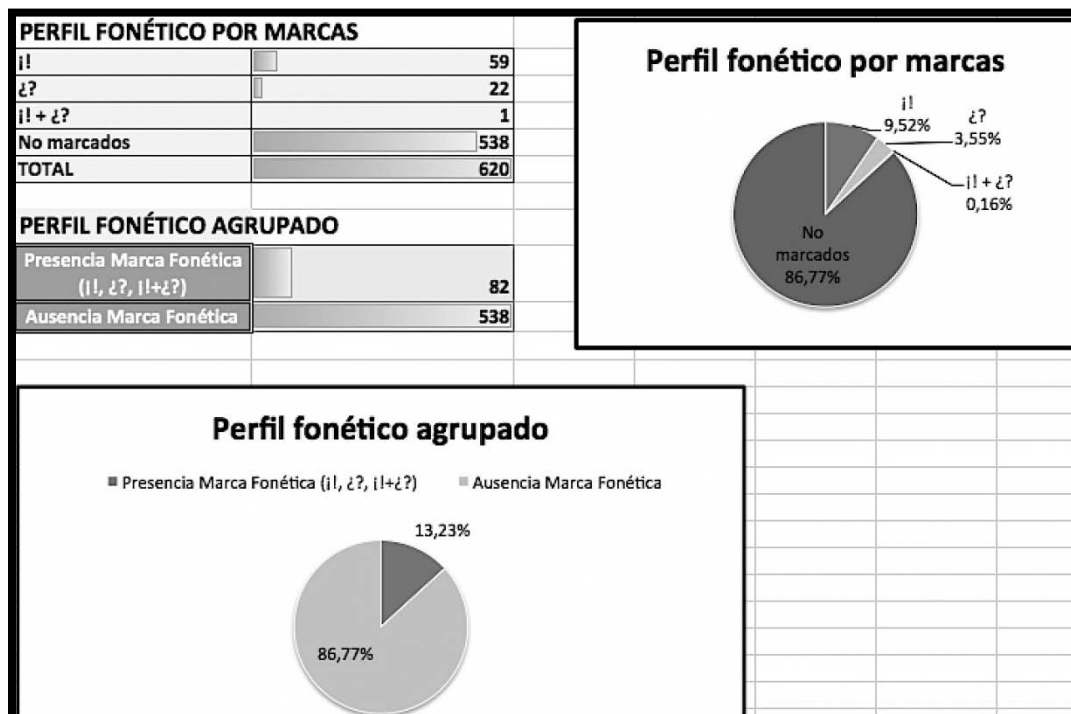
Propuesta metodológica para el estudio del lenguaje de la protesta y su  
transmisión a las redes sociales

La elección de las variables o categorías de marcación obedecerá en cualquier caso a los criterios teóricos de categorización de cada nivel. Por ejemplo, la división de las figuras retóricas (nivel 5 de análisis) en cuatro grupos corresponde a la adopción de la obra de Isabel Guerrero *Figuras Literarias y Métrica* (Catón, 1992) como guía a partir de la cual realizar la búsqueda de figuras retóricas en los elementos del corpus. Dicho manual establece una clasificación en función de la estrategia estilística, y esta clasificación es empleada para establecer los criterios de marcación o discriminación positiva. Todo establecimiento de pautas de marcación debe partir de un criterio teórico claro.

### 2.3 Marcación en función de variables de discriminación y resumen estadístico

Toda vez llevado a cabo el proceso de marcación y contabilización a través de programas útiles para el proceso (como Microsoft Excel o programas similares) de acuerdo con las variables dispuestas en el punto 2.2, se procede a condensar los resultados de dicho proceso para su uso ulterior.

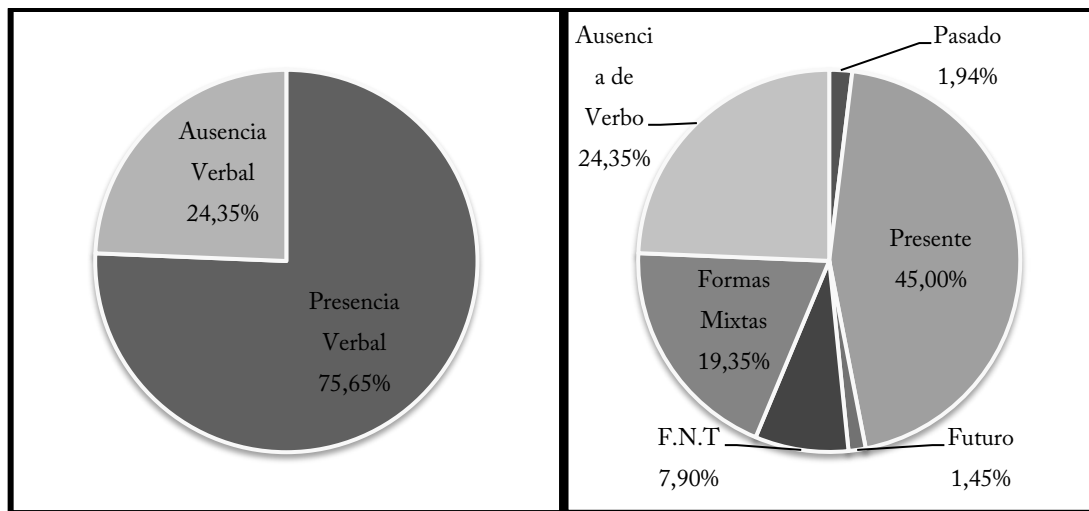
La representación gráfica del resumen de marcación de un nivel o subnivel de análisis contribuye a una mayor comprensión progresiva del estudio por parte del investigador, como se puede apreciar en la Imagen 2.3. Como protocolo adicional aconsejable, se propone disponer no sólo de los guarismos correspondientes a la marcación por categorías (en Imagen 2.3, ver “Perfil fonético por marcas”) sino en base a criterios agrupados (en la citada imagen, “Perfil fonético conjunto”).



**Imagen 2.3:** Ejemplo de resumen estadístico en archivo de marcación.  
Fuente: Martín Corvillo (2013)

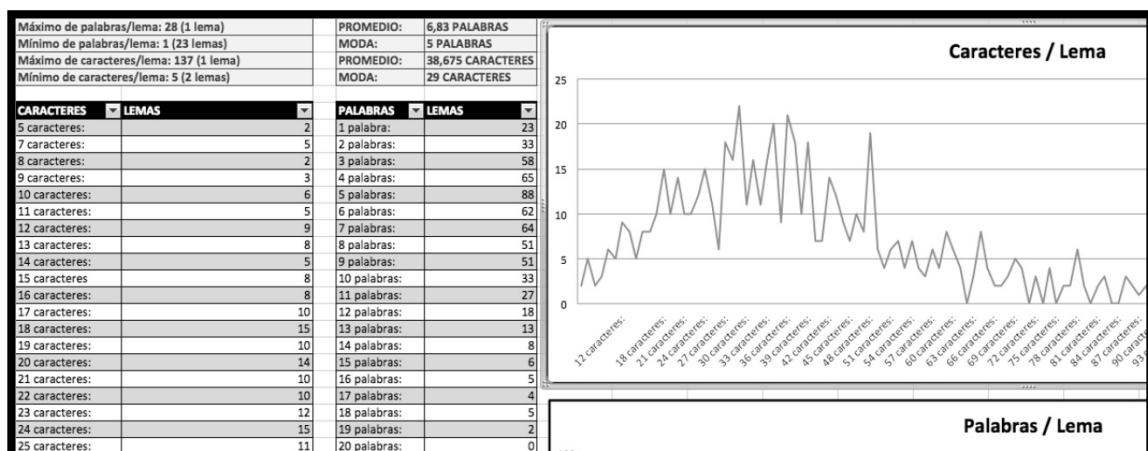
## Propuesta metodológica para el estudio del lenguaje de la protesta y su transmisión a las redes sociales

El criterio de agrupación dentro de cada nivel o subnivel de análisis, no obstante, puede variar dependiendo de las prioridades del investigador, del marco teórico o de las tendencias disciplinares del mismo. Así, si bien esto se concibe con facilidad en casos como el del perfil fonético descrito en la Imagen 2.3 –dado que el nivel de complejidad es mínimo- no se lleva a cabo sin dificultad en niveles de estudio con más variables de marcación, como el subnivel sintáctico ejemplificado en la Imagen 2.4.



**Imagen 2.4:** Ejemplo de disposición estadística agrupada (izquierda) o por marcas (derecha)  
Fuente: Martín Corvillo (2013)

En los archivos de contabilización, pese a ser en principio más sencillos dado que no nos encontraremos con variables de marcación, el nivel de detalle tenderá a ser asimismo alto. Ya se ha comentado en el punto 2.2 que, en las secciones dedicadas a la contabilización, se procuraba la disposición de un valor estadístico a priori interesante como el promedio. No obstante, esto no excluye el cálculo y disposición de valores como mediana, moda o desviación típica por citar algunos ejemplos.



**Imagen 2.5:** Ejemplo de resumen estadístico en archivo de contabilización.  
Fuente: Martín Corvillo (2013)

El objetivo principal de esta disposición estadística es facilitar al investigador la inclusión de representaciones gráficas procedentes directamente de los archivos de marcación o contabilización y no generados a posteriori mediante programas independientes de proceso de texto, como Latex o Microsoft Word. De producirse una variación en los valores estadísticos, por mínima que esta pudiese ser, los resultados de la investigación y su rigurosidad –en tanto el nivel de descripción numérica resulta tan importante de acuerdo con esta metodología- podrían verse comprometidos.

## 2.4 Establecimiento del perfil estadístico del conjunto del corpus y discriminación de las características dominantes

Tras finalizar el proceso de marcación y contabilización de acuerdo con los niveles de estudio establecidos, y tras recabar en cada nivel o subnivel los resúmenes estadísticos correspondientes, se procede a la condensación de las características dominantes en cada una de dichas parcelas de análisis.

NIVEL-SUBNIVEL	CARACTERÍSTICA DOMINANTE	PORCENTAJE	MARCAS
Fonético	Ausencia marca fonética	86,77%	538
Temporal	Presente, ausente o forma no temporal	77,25%	479
Sintáctico	Estructura Simple	56,94%	353
Temático	Economía, Política y Leyes, TIC	55,16%	342
Léxico-semántico	No parasitismo	64,68%	401
Léxico (nº palabras)	Promedio = 6,8306, Moda: 5 palabras		
Léxico (nº caracteres)	Promedio = 38,6758, Moda: 29 caracteres		
Retórico	Ausencia figuras retóricas	59,78%	370

**Tabla 2.1:** Ejemplo de disposición de características dominantes en el corpus.

Fuente: Martín Corvillo (2013)

La Tabla 2.1 ilustra las características dominantes del corpus de la tesis empleada como referencia. Con excepción de los guarismos correspondientes a las frecuencias léxicas –que resultan útiles de cara al análisis temático- todas las características preponderantes aparecen resumidas estadísticamente en este perfil conjunto. Al respecto de lo destacado como dominante en el nivel temporal, cabe señalar que se opta por aglutinar formas de presente, no temporales o directamente expresiones con ausencia verbal por resultar coincidentes en su no-referencia a eventos pasados o acontecimientos futuros.

A partir de este momento, las características dominantes serán empleadas en cualquier proceso posterior de contraste con el subcorpus.

## 2.5 Elaboración del subcorpus a través de la búsqueda de tweets mediante un buscador online

En la introducción de este artículo ya señalábamos que esta propuesta metodológica estaba orientada al análisis de la trascendencia de consignas del activismo del siglo XXI, vertidas en un contexto particular en la red. Ello condicionaba los aspectos lingüísticos a analizar y excluía factores no trasmisibles en la transición de los mensajes a la red.

Para proseguir en la tarea de deducir los factores lingüísticos que facilitan en cierto modo esa transición, el método escogido es de contraste. Así pues, toda vez se tiene un primer elemento de análisis –el corpus- es necesario obtener el segundo, al que se llamará subcorpus. Para ello, se trata de lograr el conjunto de lemas trascendidos que se ajuste más a la realidad. El medio de trascendencia escogido es Twitter, dada la relativa facilidad de rastreo de sus comunicaciones, seguramente posibilitada por la limitación estructural de las mismas a 140 caracteres.

Para la búsqueda de aquellos lemas que han podido trascender a esta red podremos emplear diversos programas gratuitos: el propio buscador de Twitter (<https://twitter.com/search-home>), SnapBird (<http://snapbird.org>) , BackTweets (<http://backtweets.com>), Twinitor (<http://www.twinitor.com>) o PostPost (<http://postpost.com>) son buenos buscadores. No obstante, su funcionamiento se centra en la cuenta de usuario de Twitter o en el seguimiento de un perfil concreto, por lo que sus resultados son limitados. Por ello, el escogido es Topsy (<http://topsy.com>) principalmente por ofrecer las siguientes posibilidades:

1. Permite ser utilizado sin tener una cuenta en Twitter.
2. Permite seguir un lema literal empleando comillas (p.ej: “Que no nos representan”) descartando resultados que no interesen.
3. Permite realizar búsquedas acotadas en el tiempo (p.ej: del 16 de mayo de 2011 al 16 de mayo de 2012).
4. Permite disponer los resultados de la búsqueda siguiendo criterios cronológicos o de influencia, lo que significa que se puede no solamente localizar un lema sino su número de replicaciones o *retweets*.
5. Permite discriminar características de los *tweets* (p.ej: si el lema lleva un enlace a una imagen que dé testimonio del propio cartel).
6. Dispone de varios idiomas, entre ellos el castellano.
7. Es gratuito.

El interfaz primario de Topsy es intuitivo como se pueda apreciar en la imagen 2.6. La introducción del texto a buscar se efectúa en la clásica barra de entrada de texto de cualquier buscador, y en la misma se puede emplear comillas para acotar la búsqueda a textos literales. Los resultados se dispondrán en un principio de acuerdo a la influencia de los *tweets* en la red social.

## Propuesta metodológica para el estudio del lenguaje de la protesta y su transmisión a las redes sociales

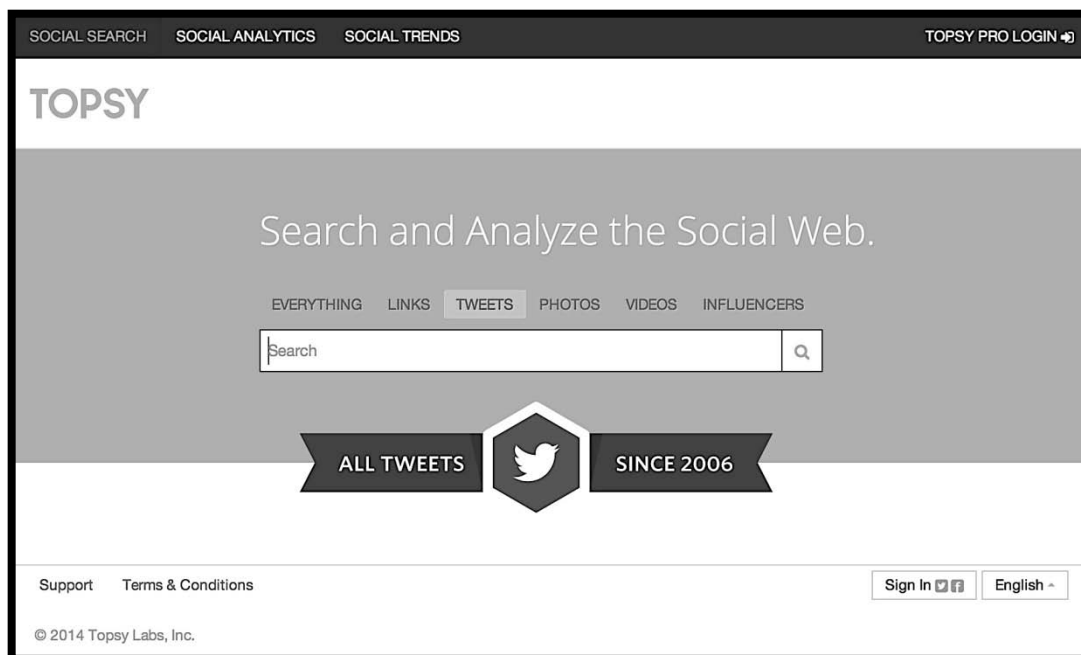


Imagen 2.6: Interfaz primario del buscador Topsy.

Fuente: <http://topsy.com>

Pese a que la influencia social en términos de alcance y replicación sea el criterio empleado por defecto por el buscador, como ya hemos indicado a la hora de enumerar las razones por las que aconsejar el uso de Topsy, los resultados pueden presentarse cronológicamente, como se aprecia en la Imagen 2.7.

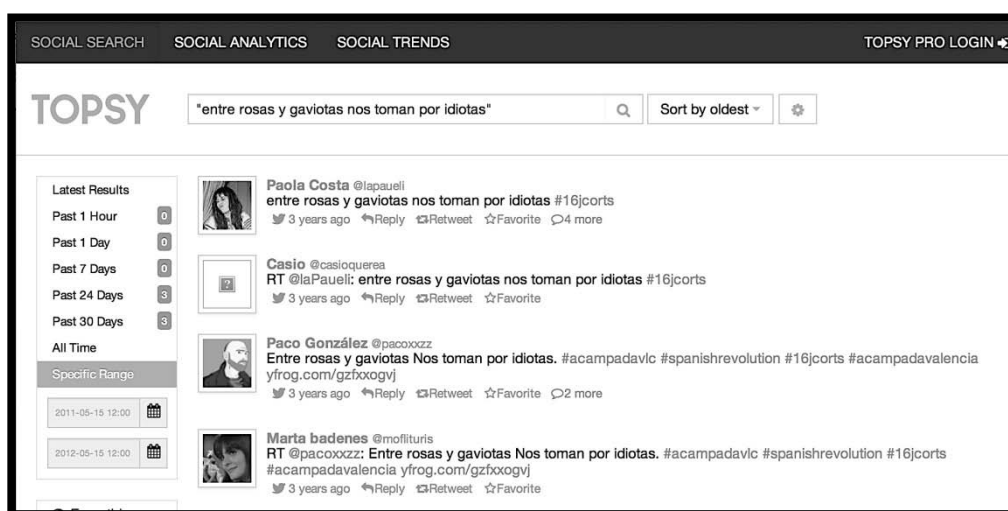


Imagen 2.7: Ejemplo de resultados arrojados por el buscador Topsy

Fuente: <http://topsy.com>



## Propuesta metodológica para el estudio del lenguaje de la protesta y su transmisión a las redes sociales

El acceso a los *tweets* es público y se realiza clicando sobre el icono de Twitter con forma de pájaro que aparece bajo el contenido del *tweet*, justo al lado de la fecha aproximada de emisión. El enlace es instantáneo y acude directamente al posteo del usuario en concreto, indicando fecha y hora de la emisión como observamos en la Imagen 2.8.



**Imagen 2.8:** Ejemplo de *tweet* con lema de corpus.

Fuente: <https://twitter.com/pacoxxzz/status/81295114441457664>

En ocasiones, la adición de un archivo fotográfico contenido en un enlace (como es el caso del *tweet* representado en la imagen 2.8) refuerza la relación de la expresión en la red social virtual con la recepción del estímulo visual del cartel original en el entorno urbano, bien sea en una concentración o manifestación, bien sea al circular por uno de los espacios ocupados por el movimiento de protesta. La presencia de un testimonio visual como el representado en la Imagen 2.9 refuerza la tesis de la relación entre la comunicación visual de un movimiento de protesta y la existencia de redes sociales.



**Imagen 2.9:** Ejemplo de imagen de cartel enlazado en *tweet*.

Fuente: <http://yfrog.com/gzfxogvj>



## Propuesta metodológica para el estudio del lenguaje de la protesta y su transmisión a las redes sociales

El procedimiento de búsqueda se complementa mediante un procedimiento de marcación cuya variable de discriminación es “trascendido”. Con el fin de agilizar el proceso, será conveniente registrar en la misma tabla datos de futura utilidad, añadiendo las columnas de marcación pertinentes de ser necesario. Cuestiones como el contenido escrito del *tweet*, el *link* o la fecha parecen esenciales para dar veracidad a la investigación. Otras cuestiones de diferenciación, como el grado de literalidad de la trascendencia, el número de *retweets* – expuestas en la imagen 2.10- o si se acompaña de una imagen, pueden ser cuestiones a tener en cuenta.

LEMA	TRASCENDENCIA		CONTENIDO DEL TWEET	LINK	FECHA (DD/MM/AAAA)
	INTRASCENDENTE (NO TWITTEADO)	FORMAL (TWITTEADO LITERAL/TRADUCCIÓN/HASHTAG)			
A la puta calle nos dijeron y eso hacemos					
A por ellos, que son pocos y cobardes					
¡A por la tercera ya!					
Abre los ojos. Estamos aquí.					
Abuelo esto va x ti					
Ahí está la cueva de Alí Babá		1	'Ahí está la cueva de Ali Baba' <a href="http://bit.ly/m89l8u#19j">http://bit.ly/m89l8u#19j</a> #acampadagranada	<a href="https://twitter.com/#!/PedroUve/status/82654792039677952">https://twitter.com/#!/PedroUve/status/82654792039677952</a>	20/06/11
Ahora o nunca		1	"Ahora o nunca, ahora o nunca"	<a href="https://twitter.com/#!/hugomabarc/status/70567881959673856">https://twitter.com/#!/hugomabarc/status/70567881959673856</a>	17/05/11

Imagen 2.10: Ejemplo de marcación y adición de información de los lemas trascendidos.

Fuente: Martín Corvillo (2013)

Finalmente, el conjunto de lemas del corpus que han trascendido a la red es aislado y conforma el subcorpus. Se ha de recordar la conveniencia de no volver a numerar los lemas del subcorpus como si se tratase de un corpus totalmente independiente. Si bien los procesos de análisis posteriores se realizarán exclusivamente en esa sección del corpus, el mantenimiento de las referencias fijadas al principio de la investigación facilita la identificación y revisión de cada lema en cualquier estadio del proceso de investigación.

Toda vez seleccionado el subcorpus del corpus original, se procede al aislamiento del mismo. Una estrategia que puede facilitar este procedimiento –y que se ve facilitada por la comentada fijación de una misma referencia para cada elemento del corpus- es la ubicación en cada archivo de marcación del corpus de una columna exclusiva de discriminación de lemas trascendidos. Eso dota al archivo de la posibilidad de excluir los lemas no marcados sin perder la información de cada nivel o subnivel de análisis.

## 2.6 Establecimiento del perfil estadístico del conjunto del subcorpus y discriminación del estado de las características dominantes del corpus

La citada estrategia de discriminación del subcorpus a partir de la variable de marcación “trascendido” genera automáticamente los resúmenes estadísticos parciales del subcorpus, facilitando el establecimiento del perfil estadístico del conjunto del subcorpus cuya ejemplificación podemos advertir en la Tabla 2.2.

NIVEL -SUBNIVEL	CARACTERÍSTICA DOMINANTE	PORCENTAJE	MARCAS
Fonético	Ausencia marca fonética	88,86%	335
Temporal	Presente, ausente o forma no temporal	78,78%	297
Sintáctico	Estructura Simple	59,15%	223
Temático	Economía, Política y Leyes, TIC	55,97%	211
Léxico-semántico	No parasitismo	56,23%	212
Léxico (nº palabras)	Promedio= 6,46; Moda: 5 palabras		
Léxico (nº caracteres)	Promedio=36,1; Moda: 37 caracteres		
Retórico	Ausencia figuras retóricas	59,42%	224

**Tabla 2.2:** Ejemplo de disposición de características dominantes en el subcorpus.  
Fuente: Martín Corvillo (2013)

El establecimiento del perfil estadístico del conjunto del subcorpus, no obstante, se realiza a partir de las características dominantes del corpus. Esto es: se disponen los valores actualizados de las mismas predominancias que en el conjunto completo de lemas, sustituyendo los valores porcentuales, número de marcas y demás valores estadísticos que hayan sido tomados en cuenta en el perfil estadístico del conjunto del corpus.

## 2.7 Contraste entre los perfiles estadísticos conjuntos de corpus y subcorpus

A la hora de realizar una comparativa entre los perfiles estadísticos conjuntos de corpus y subcorpus, es habitual que los parámetros que eran predominantes en el corpus inicial dejen de serlo en el mismo grado o por el contrario se potencien, variando su presencia a la baja o al alza. Variaciones sensibles en términos porcentuales son indicativas de la importancia de los parámetros continentales en términos de trascendencia. Así, considerando la variación entre los valores reflejados en la Tabla 2.1 y la Tabla 2.2, apreciamos cómo los parámetros “Estructura simple” del nivel sintáctico de análisis y “No parasitismo” en el nivel léxico-semántico ofrecen fluctuaciones sensiblemente superiores al del resto de valores de los conjuntos. “Estructura simple” y “No parasitismo” fluctúan de manera evidente. En este caso, tanto el uso de estructuras sintácticas simples como el uso de lemas parasitarios –“No parasitismo” desciende, con lo que su parámetro opuesto asciende por el mismo valor- son determinantes a la hora de explicar las características que posibilitan que un lema sea más susceptible de traspasar la barrera entre la comunicación visual del entorno urbano y el virtual.

### 3. Conclusiones

Como se expuso en la introducción, el objetivo de este artículo es dotar a los investigadores del ámbito de la lingüística de corpus de una metodología de estudio que pueda servir para analizar las comunicaciones de un movimiento social del siglo XXI. Es decir, una propuesta metodológica adaptable a las nuevas vías de comunicación mediante las cuales se lleva a cabo la coordinación de los movimientos sociales, pero también la propagación de sus lemas.

La compleja realidad actual de los movimientos sociales, inmersos en un proceso de hibridación hacia un nuevo medio no presencial, supone un estímulo a la investigación no sólo sociológica sino también lingüística. El desarrollo de propuestas metodológicas de investigación que no renuncien ni a un alto nivel descriptivo ni al valor de las descripciones estadísticas se antoja necesario.

**José Manuel Martín Corvillo**

Doctor por la Universitat de València

[jomarcor@alumni.uv.es](mailto:jomarcor@alumni.uv.es)

## Referencias bibliográficas

- Della Porta, D. y Diani, M. (2011, 1.ª ed.): *Los Movimientos Sociales*, Madrid: Centro de Investigaciones Sociológicas: Editorial Complutense.
- Guerrero, I. (1992, 1.ª ed.): *Figuras literarias y métrica*, Zaragoza: Catón.
- Martín Corvillo, J. M. (2013): *15M: Análisis de la entropía comunicativa* (Tesis doctoral inédita), Universitat de Valencia. España.
- Sperber, D. y Wilson, D. (2004): "La teoría de la Relevancia", *Revista de Investigación Lingüística*, 7, pp. 237-286.
- Tilly, C. y Wood, L. J. (2009, 1.ª ed.): *Los movimientos sociales, 1768-2008: desde sus orígenes a Facebook*, Barcelona: Crítica.
- Torres Nabel, L. C. (2009): "Redes sociales y efectos políticos. Reflexiones sobre el impacto de twitter México. Sociología y tecnociencia", *Revista Digital de Sociología del Sistema Tecnocientífico*, 0 (2), pp. 63-75.
- Yus, F. (2010, 1.ª ed.): *Ciberpragmática 2.0. Nuevos usos del lenguaje en Internet*, Barcelona: Ariel.